MA2001 Notes

Homura Akemi

NUS AY2021-2022 Sem 2

1 Linear Systems and Gaussian Elimination

1.1 Linear Systems and Their Solutions

- A **linear equation** in n variables $x_1, x_2, \ldots, x_n$ is an equation of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

  where $a_1, a_2, \ldots, a_n, b \in \mathbb{R}$
  - If $a_1 = \cdots = a_n = 0$ but $b \neq 0$, then it is **inconsistent**
  - If $a_1 = \cdots = a_n = b = 0$, then it is a **zero equation**
  - A linear eq that is not zero is a **nonzero equation**
- For $s_1, \ldots, s_n \in \mathbb{R}$, $x_1 = s_1, \ldots, x_n = s_n$ is a **solution** if

$$a_1s_1 + a_2s_2 + \cdots + a_ns_n = b$$

- The set of all solutions is the **solution set**
- An expression that gives the entire solution set is a **general solution**
- A **linear system (system of linear equations)** of m linear equations in n variables $x_1, x_2, \ldots, x_n$ is

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

  where $a_{ij}, b_i \in \mathbb{R}$
- **Solution**, **solution set** and **general solution** are defined analogously for a linear system
- A linear system is **consistent** if it has at least one solution and **inconsistent** if it has none

1.2 Elementary Row Operations

- For a given linear system such as the one above, its **augmented matrix** is

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{array}\right)$$

- The following operations on rows of an augmented matrix are called **elementary row operations**
  - Multiply a row by a nonzero constant
  - Interchange two rows
  - Add a constant multiple of a row to another
- Two augmented matrices are **row equivalent** if one can be obtained from the other via EROs
- **Theorem** - If two augmented matrices are row equivalent, then their systems have the same solution

1.3 Row-Echelon Form

- An augmented matrix is in **row-echelon form** if
  - The zero rows are grouped together at the bottom
  - For any two successive nonzero rows, the first nonzero number (**leading entry**) in the lower row appears to the right of the first nonzero number in the higher row
- Suppose a matrix is in REF
  - The leading entry of a nonzero row is a **pivot point**
  - A column is a **pivot column** if it has a pivot point and **non-pivot column** if not
- An augmented matrix in REF is in **reduced row-echelon form** if
  - The leading entry of every nonzero row is 1
  - In every pivot column, all entries except the pivot point are 0
- Suppose the augmented matrix of a linear system is in REF
  - Set the variables of non-pivot columns to arbitrary parameters
  - Solve the variables of pivot columns by back substitution (bottom to top)

1.4 Gaussian Elimination

- Gaussian Elimination (for REF)
  - Find the leftmost column that is not all 0
  - Check the top entry of the column. If it's 0, make it nonzero by swapping rows
  - For each row below the top row, add a multiple of the top row to make the rest of the column 0
  - Cover the top row and repeat until done
- Gauss-Jordan Elimination (for RREF)
  - Do Gaussian Elimination
  - Multiply rows by constants to make all pivot points 1
  - Starting from the last row, add a multiple of it to other rows to make the rest of the pivot column 0
- Consistency
  - No solution - last column is a pivot column
  - 1 solution - only the last column is a non-pivot column
  - Infinite solutions - last column and at least one other column are non-pivot columns
  - * Number of parameters = number of non-pivot columns except the last
- Notations for EROs
  - kR_i - multiply R_i row by k
  - $R_i \leftrightarrow R_j$ - swap R_i and R_j
  - $R_j + kR_i$ - add kR_i to R_j

1.5 Homogeneous Linear Systems

- A linear equation in $x_1, x_2, \dots, x_n$ is called **homogeneous** if

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0 \iff x_1 = x_2 = \cdots = x_n = 0 \text{ is a solution}$$

- A linear system is **homogeneous** if every equation is homogeneous
- $x_1 = x_2 = \cdots = x_n = 0$ is the **trivial solution** of a homogeneous linear system, while all other solutions are **non-trivial solutions**

2 Matrices

2.1 Introduction to Matrices

- A **matrix** is a rectangular array of numbers. The matrix below has m rows and n columns.

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

- The **size** of the matrix is $m \times n$
- The **(i, j)-entry** (a_{ij}) is the entry in the ith row and jth column
- It is denoted by $\boldsymbol{A} = (a_{ij})_{m \times n}$, or just $\boldsymbol{A} = (a_{ij})$ if the size is known/not important
- A **row/column matrix (row/column vector)** is a matrix with only one row/column
- A **square matrix** is a matrix with the same number of rows and columns
- An $n \times n$ square matrix has **order** n
- The **diagonal** of a square matrix $\boldsymbol{A}$ is the sequence of entries $a_{11}, a_{22}, \dots, a_{nn}$
- a_{ij} is $\begin{cases} \text{a } \textbf{diagonal entry} & i = j \\ \text{a } \textbf{non-diagonal entry} & i \neq j \end{cases}$
- **Diagonal matrix** - square matrix where all non-diagonal entries are 0
- **Scalar matrix** - diagonal matrix with all equal diagonal entries
- **Identity matrix** - scalar matrix with all diagonal entries 1
- **Zero matrix** - matrix with all entries 0
- **Symmetric matrix** - square matrix that is symmetric wrt the diagonal
- **Upper/Lower triangular matrix** - square matrix with all entries below/above the diagonal 0

2.2 Matrix Operations

- Let $\boldsymbol{A} = (a_{ij})_{m \times p}, \boldsymbol{B} = (b_{ij})_{p \times n}$. Then $\boldsymbol{AB}$ is the $m \times n$ matrix whose (i, j)-entry is

$$\sum_{k=1}^{p} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{ip}b_{pj}$$

- Matrix multiplication is **not commutative**
- $\boldsymbol{AB}$ is the pre-multiplication of $\mathbf{A}$ to $\mathbf{B}$ and post-multiplication of $\mathbf{B}$ to $\mathbf{A}$
- **Properties of Matrix Multiplication**
  - $\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C}$
  - $\boldsymbol{A}(\boldsymbol{B}_1 + \boldsymbol{B}_2) = \boldsymbol{AB}_1 + \boldsymbol{AB}_2$ and $(\boldsymbol{A}_1 + \boldsymbol{A}_2)\boldsymbol{B} = \boldsymbol{A}_1\boldsymbol{B} + \boldsymbol{A}_2\boldsymbol{B}$
  - $c(\boldsymbol{AB}) = (c\boldsymbol{A})\boldsymbol{B} = \boldsymbol{A}(c\boldsymbol{B})$
  - $\boldsymbol{A0} = \boldsymbol{0A} = \boldsymbol{0}$
  - $\boldsymbol{AI} = \boldsymbol{IA} = \boldsymbol{A}$

- Let $\boldsymbol{A}$ be a square matrix of order n. It's **powers** are defined as

$$\boldsymbol{A}^k = \begin{cases} \boldsymbol{I}_n & \text{if } k = 0 \\ \underbrace{\boldsymbol{A}\boldsymbol{A}\dots\boldsymbol{A}}_{k \text{ times}} & \text{if } k \geq 1 \end{cases}$$

- Let $\boldsymbol{A} = (a_{ij})_{m\times n}$ and $a_i = \begin{pmatrix} a_{i1} & a_{i2} & \dots a_{in} \end{pmatrix}$ denote the ith row. Then $\boldsymbol{A} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}$.

- Similarly, if b_j is the jth column, then $\boldsymbol{A} = \begin{pmatrix} b_1 & b_2 & \dots & b_n \end{pmatrix}$

- If $\boldsymbol{A} = (a_{ij})_{m\times p}$ with ith row a_i and $\boldsymbol{B} = (a_{ij})_{p\times n}$ with jth column b_j, then

$$\boldsymbol{AB} = \begin{pmatrix} a_1b_1 & a_1b_2 & \dots & a_1b_n \\ a_2b_1 & a_2b_2 & \dots & a_2b_n \\ \vdots & \vdots & & \vdots \\ a_mb_1 & a_mb_2 & \dots & a_mb_n \end{pmatrix} = \begin{pmatrix} a_1\boldsymbol{B} \\ a_2\boldsymbol{B} \\ \vdots \\ a_m\boldsymbol{B} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}b_1 & \boldsymbol{A}b_2 & \dots & \boldsymbol{A}b_n \end{pmatrix}$$

- Consider the linear system

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \dots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

  - Coefficient matrix: $\boldsymbol{A} = (a_{ij})_{m\times n}$
  - Variable matrix: $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$
  - Constant matrix: $\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$
  - $\boldsymbol{Ax} = \boldsymbol{b}$

- Linear system with > 1 solution $\Rightarrow$ infinite solutions
  Sketch of Proof: If $Ax = b$ has two distinct solutions u_1, u_2, then $u_2 + t(u_1 - u_2)$ is a solution $\forall t \in \mathbb{R}$

- The **transpose** of a matrix $\boldsymbol{A} = (a_{ij})_{m\times n}$ is the $n \times m$ matrix $\boldsymbol{A}^T$ whose (i, j)-entry is a_{ji}

- **Theorem - Properties of Transpose**
  - $(\boldsymbol{A}^T)^T = \boldsymbol{A}$
  - $\boldsymbol{A}$ is symmetric $\iff \boldsymbol{A} = \boldsymbol{A}^T$
  - $(c\boldsymbol{A})^T = c\boldsymbol{A}^T$
  - $(\boldsymbol{A} + \boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$
  - $(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$

2.3 Inverses of Square Matrices

- If $\boldsymbol{A}$ is a square matrix of order n and $\exists$ $\boldsymbol{B}$ st $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I_n}$, then $\boldsymbol{A}$ is **invertible** and $\boldsymbol{B}$ is an **inverse** of $\boldsymbol{A}$
- If $\boldsymbol{A}$ is not invertible, then it is **singular**

* Non-square matrices are neither invertible nor singular

- If $\boldsymbol{A}$ is invertible, it's inverse is unique
- **Cancellation Law** - If $\boldsymbol{A}$ is invertible, then
  - $\boldsymbol{AB}_1 = \boldsymbol{AB}_2 \Rightarrow \boldsymbol{B}_1 = \boldsymbol{B}_2$
  - $\boldsymbol{C}_1\boldsymbol{A} = \boldsymbol{C}_2\boldsymbol{A} \Rightarrow \boldsymbol{C}_1 = \boldsymbol{C}_2$
- **Theorem - Properties of Inverses**
  - $(c\boldsymbol{A})^{-1} = \frac{1}{c}\boldsymbol{A}^{-1}$
  - $(\boldsymbol{A}^T)^{-1} = (\boldsymbol{A}^{-1})^T$
  - $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$
  - $(\boldsymbol{A}_1\boldsymbol{A}_2 \dots \boldsymbol{A}_k)^{-1} = \boldsymbol{A}_k^{-1} \dots \boldsymbol{A}_2^{-1}\boldsymbol{A}_1^{-1}$ (by induction)
  - $\boldsymbol{A}^{-k} = (\boldsymbol{A}^{-1})^k = (\boldsymbol{A}^k)^{-1}$

2.4 Elementary Matrices

- A square matrix is an **elementary matrix** if it can be obtained from the identity matrix by performing a single elementary row operation
- **Theorem** - If $\boldsymbol{E}$ is the elementary matrix obtained by performing an ERO to $\boldsymbol{I}_m$, then for any $m \times n$ matrix $\boldsymbol{A}$, $\boldsymbol{EA}$ can be obtained by performing the same ERO to $\boldsymbol{A}$
- **Theorem** - Every elementary matrix has an inverse that is also elementary
- **Theorem** - $\boldsymbol{A}$ and $\boldsymbol{B}$ are row equiv $\iff$ $\exists$ elementary matrices $\boldsymbol{E}_1, \dots, \boldsymbol{E}_k$ such that $\boldsymbol{B} = \boldsymbol{E}_k\boldsymbol{E}_{k-1} \dots \boldsymbol{E}_1\boldsymbol{A}$
- **Theorem** - Augmented matrices of two linear systems are row equivalent $\Rightarrow$ same solution set
  Sketch of Proof: Let them be $\boldsymbol{Ax} = \boldsymbol{b}, \boldsymbol{Cx} = \boldsymbol{d}$. Perform EROs to go from one to the other.
- **Main Theorem for Invertible Matrices** - Let $\boldsymbol{A}$ be a square matrix. The ff are equivalent
  1. $\boldsymbol{A}$ is an invertible matrix
  2. Linear system $\boldsymbol{Ax} = \boldsymbol{b}$ has a unique solution $\forall$ $\boldsymbol{b}$
  3. Linear system $\boldsymbol{Ax} = \boldsymbol{0}$ only has the trivial solution
  4. The RREF of $\boldsymbol{A}$ is $\boldsymbol{I}$
  5. $\boldsymbol{A}$ is the product of elementary matrices

  Sketch of Proof: Show $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 5 \Rightarrow 1$
- **Theorem** - Let $\boldsymbol{A}$ be an invertible matrix. The RREF of $(\boldsymbol{A} \mid \boldsymbol{I})$ is $(\boldsymbol{I} \mid \boldsymbol{A}^{-1})$
- **Theorem** - Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be square matrices of the same size. If $\boldsymbol{AB} = \boldsymbol{I}$, then $\boldsymbol{A}$ and $\boldsymbol{B}$ are invertible and $\boldsymbol{A}^{-1} = \boldsymbol{B}, \boldsymbol{B}^{-1} = \boldsymbol{A}$
  Sketch of Proof: $\boldsymbol{Bx} = \boldsymbol{0} \Rightarrow \boldsymbol{x} = \boldsymbol{ABx} = \boldsymbol{A0} = \boldsymbol{0}$ only solution $\Rightarrow \boldsymbol{B}$ invertible then use Cancellation Law
- **Corollary** - Let $\boldsymbol{A}_1, \boldsymbol{A}_2, \dots, \boldsymbol{A}_k$ be square matrices of the same size. Then
  $$\boldsymbol{A}_1, \boldsymbol{A}_2, \dots, \boldsymbol{A}_k \text{ is invertible} \iff \text{all } \boldsymbol{A}_i \text{ are invertible}$$
- **Elementary Column Operations** - EROs but on columns
- If $\boldsymbol{E}$ is obtained from $\boldsymbol{I}_n$ by a single elementary column operation, then $\boldsymbol{E}$ is an elementary matrix
- If $\boldsymbol{E}$ is the elementary matrix obtained by performing an ECO to $\boldsymbol{I}_n$, then for any $m \times n$ matrix $\boldsymbol{A}$, $\boldsymbol{AE}$ can be obtained by performing the same ECO to $\boldsymbol{A}$

2.5 Determinant

- Let $\boldsymbol{M}_{ij}$ be the submatrix of $\boldsymbol{A}$ from deleting the ith row and jth column. Then the **(i,j)-cofactor** of $\boldsymbol{A}$ is
$$A_{ij} = (-1)^{i+j}\det(\boldsymbol{M}_{ij})$$
- Let $\boldsymbol{A} = (a_{ij})_{n\times n}$ and A_{ij} be its (i,j)-cofactor. Then its **determinant** is
$$\det(\boldsymbol{A}) = \begin{cases} a_{11} & n = 1 \\ a_{11}A_{11} + a_{12}A_{12} + \cdots + a_{1n}A_{1n} & n > 1 \end{cases}$$
- **Theorem - Determinants Under EROs**
  - $\boldsymbol{A} \xrightarrow{cR_i} \boldsymbol{B} \Rightarrow \det(\boldsymbol{B}) = c\ \det(\boldsymbol{A})$
  - $\boldsymbol{A} \xrightarrow{R_i \leftrightarrow R_j} \boldsymbol{B} \Rightarrow \det(\boldsymbol{B}) = -\det(\boldsymbol{A})$
  - $\boldsymbol{A} \xrightarrow{R_i + cR_j} \boldsymbol{B} \Rightarrow \det(\boldsymbol{B}) = \det(\boldsymbol{A})$
- **Theorem** - Let $\boldsymbol{A}$ be a square matrix. For any elementary matrix $\boldsymbol{E}$, $\det(\boldsymbol{EA}) = \det(\boldsymbol{E})\det(\boldsymbol{A})$
- **Theorem** - If square matrix $\boldsymbol{A}$ has a zero row, then $\det(\boldsymbol{A}) = 0$
- If matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are row equivalent, then $\det(\boldsymbol{A}) = 0 \iff \det(\boldsymbol{B}) = 0$
- **Theorem** - $\det(\boldsymbol{A}) = 0 \iff \boldsymbol{A}$ is singular / $\det(\boldsymbol{A}) \neq 0 \iff \boldsymbol{A}$ is invertible
  Sketch of Proof: Use the fact that the RREF of $\boldsymbol{A}$ is $\boldsymbol{I}$ iff it's invertible
- **Theorem** - $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$
  Sketch of Proof: Split $\boldsymbol{A}$ into a product of elementary matrices
- **Theorem** - $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^T)$
  Sketch of Proof: Split $\boldsymbol{A}$ into elementary matrices and note that $\det(\boldsymbol{E}) = \det(\boldsymbol{E}^T)$
- **Theorem** - If $\boldsymbol{A} = (a_{ij})_{n\times n}$ is triangular, then $\det(\boldsymbol{A}) = a_{11}a_{22}\dots a_{nn}$
  Sketch of Proof: WLOG upper-triangular. Consider the elementary matrices used to bring it to its RREF $\boldsymbol{I}$
- **Theorem (Cofactor Expansion)** - Let $\boldsymbol{A}$ be a matrix of order n and A_{ij} be its (i,j)-cofactor. Then $\forall\ i, j$,
$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} a_{ik}A_{ik} = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in}$$
$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} a_{kj}A_{kj} = a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj}$$
  Sketch of Proof:
  - Let $\boldsymbol{B} = (b_{ij})$ be the matrix obtained by moving the ith row of $\boldsymbol{A}$ to the top for some i
  - This can be done by $i-1$ swaps, so $\det(\boldsymbol{A}) = (-1)^{i-1}\det(\boldsymbol{B})$
  - Note that matrix from removing the 1st row and jth column of $\boldsymbol{B}$ is identical to the matrix from removing the ith row and jth column of $\boldsymbol{A}$, so $B_{ij} = (-1)^{i-1}A_{ij}$
  - $\therefore\ \det(\boldsymbol{A}) = (-1)^{i-1}\det(\boldsymbol{B}) = (-1)^{i-1}\sum_{k=1}^{n} b_{1k}B_{1k} = (-1)^{i-1}\sum_{k=1}^{n} a_{ik}(-1)^{i-1}A_{ik} = \sum_{k=1}^{n} a_{ik}A_{ik}$
- **Determinant Formulas** (last 2 are exercises)
  - $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^T)$
  - $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$
  - $\det(c\boldsymbol{A}) = c^n\det(\boldsymbol{A})$
  - $\det(\boldsymbol{A}^{-1}) = \det(\boldsymbol{A})^{-1}$ if $\boldsymbol{A}$ is invertible

- Let $\boldsymbol{A}$ be a square matrix of order n. The **(classical) adjoint** (or **adjugate** or **adjunct**) of $\boldsymbol{A}$ is

$$\mathbf{adj}(\boldsymbol{A}) = (A_{ji})_{n\times n}$$

  where A_{ij} is the (i,j)-cofactor of $\boldsymbol{A}$

- **Theorem** - Let $\boldsymbol{A}$ be a square matrix. Then $\boldsymbol{A}[\mathbf{adj}(\boldsymbol{A})] = \det(\boldsymbol{A})\boldsymbol{I}$

  Proof: Let $\boldsymbol{A}[\mathbf{adj}(\boldsymbol{A})] = (c_{ij})$. Then $c_{ii} = \displaystyle\sum_{k=1}^{n} a_{in}A_{in} = \det(\boldsymbol{A})$.

  If $i \neq j$, let $\boldsymbol{B}$ be $\boldsymbol{A}$ with jth row replaced by ith row. Then $c_{ij} = \displaystyle\sum_{k=1}^{n} a_{ik}A_{jk} = \sum_{k=1}^{n} b_{jk}B_{jk} = \det(\boldsymbol{B}) = 0$

- **Theorem** - Let $\boldsymbol{A}$ be a square matrix. Then $\boldsymbol{A}[\mathbf{adj}(\boldsymbol{A})] = [\mathbf{adj}(\boldsymbol{A})]\boldsymbol{A} = \det(\boldsymbol{A})\boldsymbol{I}$, and if $\boldsymbol{A}$ is invertible, then $\boldsymbol{A}^{-1} = \dfrac{1}{\det(\boldsymbol{A})}\mathbf{adj}(\boldsymbol{A})$

- **Cramer's Rule** - Let $\boldsymbol{A}$ be an invertible matrix of order n. Then for every column matrix $\boldsymbol{b}$ of size $n \times 1$, the linear system $\boldsymbol{Ax} = \boldsymbol{b}$ has the unique solution

$$\boldsymbol{x} = \frac{1}{\det(\boldsymbol{A})}\begin{pmatrix} \det(\boldsymbol{A}_1) \\ \det(\boldsymbol{A}_2) \\ \vdots \\ \det(\boldsymbol{A}_n) \end{pmatrix}$$

  where $\boldsymbol{A}_j$ is obtained by replacing the jth column of $\boldsymbol{A}$ with $\boldsymbol{b}$

3 Vector Spaces

3.1 Euclidean n-Spaces

- An **n-vector** or **ordered n-tuple** of real numbers is $\boldsymbol{v} = (v_1, v_2, \dots, v_n)$
- v_i is the **ith component** or **ith coordinate** of $\boldsymbol{v}$
- Let $\boldsymbol{u} = (u_1, u_2, \dots, u_n)$ and $\boldsymbol{v} = (v_1, v_2, \dots, v_n)$. Then $\boldsymbol{u} = \boldsymbol{v} \iff u_i = v_i \ \forall\ i = 1, \dots, n$
- The n-vector $(0, 0, \dots, 0)$ is the **zero vector**
- Let $c \in \mathbb{R}$. The **scalar multiple** $c\boldsymbol{v}$ is $(cv_1, cv_2, \dots, cv_n)$
- An n-vector can be viewed as a row or column matrix
- The **Euclidean n-space** (or simply **n-space**) is the set of all n-vectors of real numbers

$$\mathbb{R}^n = \{(v_1, v_2, \dots, v_n) | v_1, v_2, \dots, v_n \in \mathbb{R}\}$$

- Solution set to linear system $\boldsymbol{Ax} = \boldsymbol{b}$ with n variables is a subset of $\mathbb{R}^n$
- A linear system is given in the **implicit form** and its general solution is in the **explicit form**

3.2 Linear Combinations and Linear Spans

- Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k$ be vectors. $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k$, is a **linear combination** of them, where $c_1, c_2, \dots, c_k \in \mathbb{R}$
- Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$ be a subset of $\mathbb{R}^n$. The set of all linear combinations of $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k$

$$\{c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k | c_1, c_2, \dots, c_k \in \mathbb{R}\}$$

  is called the **linear span** of S, denoted by $\text{span}(S)$ or $\text{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$

- Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\} \subseteq \mathbb{R}^n$. View each $\boldsymbol{v}_i$ as a column vector and let $\boldsymbol{A} = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \ldots\ \boldsymbol{v}_k)$. Then
$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k = \boldsymbol{A}\begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix}$$
so
$$\text{span}(S) = \mathbb{R}^n \iff \boldsymbol{Ax} = \boldsymbol{v} \text{ consistent } \forall\ \boldsymbol{v} \in \mathbb{R}^n \iff \text{REF of } \boldsymbol{A} \text{ has no zero row}$$
- **Theorem** - Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\} \subseteq \mathbb{R}^n$. If $k < n$, then $\text{span}(S) \neq \mathbb{R}^n$
- **Theorem** - $\boldsymbol{0} \in \text{span}(S)$ for any $S \subseteq \mathbb{R}^n$, where $\boldsymbol{0}$ is the zero vector in $\mathbb{R}^n$
- **Theorem** - Given two subsets of $\mathbb{R}^n$, $S_1 = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$, $S_2 = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$,
$$\text{span}(S_1) \subseteq \text{span}(S_2) \iff \text{Every } \boldsymbol{u}_i \text{ is a linear combination of } \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$$
- **Theorem** - If $\boldsymbol{v}_k$ is a linear combination of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{k-1}$, then $\text{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{k-1}\} = \text{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{k-1}, \boldsymbol{v}_k\}$
- Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\} \subseteq \mathbb{R}^n$.
$$\boldsymbol{v} \in \text{span}(S) \iff \boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k \text{ for some } c_i \in \mathbb{R} \iff (\boldsymbol{v}_1\ \ldots\ \boldsymbol{v}_k)\begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix} = \boldsymbol{v}$$
View each $\boldsymbol{v}_i$ as a column vector and let $\boldsymbol{A} = (\boldsymbol{v}_1\ \ldots\ \boldsymbol{v}_k)$. Then $\boldsymbol{Ax} = \boldsymbol{v}$ is consistent $\iff \boldsymbol{v} \in \text{span}(S)$

3.3 Subspaces

- Let V be a subset of $\mathbb{R}^n$. Then V is a **subspace** of $\mathbb{R}^n$ if $\exists\ \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$ st $V = \text{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$
- V is the **subspace spanned** by $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$
- $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ **spans** the subspace V
- Let $\boldsymbol{0} \in \mathbb{R}^n$ be the zero vector. Then $\{\boldsymbol{0}\} = \text{span}\{\boldsymbol{0}\}$ is the **zero space**
- Since a subspace V is of the form $\text{span}(S)$, then
  - $\boldsymbol{0} \in V$
  - $c \in \mathbb{R}$ and $\boldsymbol{v} \in V \Rightarrow c\boldsymbol{v} \in V$
  - $\boldsymbol{u}, \boldsymbol{v} \in V \Rightarrow \boldsymbol{u} + \boldsymbol{v} \in V$

  If any of the above fails, then V is not a subspace (of $\mathbb{R}^n$)
- **Theorem** - The **solution set** of a homogeneous linear system of n variables is a subspace of $\mathbb{R}^n$
Sketch of Proof: Solve the homogeneous system using its RREF in terms of the arbitrary parameters
- The solution set of a homogeneous linear system is called the **solution space** of the system

3.4 Linear Independence

- Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\} \subseteq \mathbb{R}^n$. The equation
$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k = \boldsymbol{0}$$
has a **trivial solution** $c_1 = c_2 = \cdots = c_k = 0$
- If the equation has a non-trivial solution, then
  - S is a **linearly dependent set**
  - $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ are **linearly dependent**

  - $\exists\ c_1, c_2, \dots c_k \in \mathbb{R}$ not all zero st $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k = \mathbf{0}$

- If the equation only has the trivial solution, then
  - S is a **linearly independent set**
  - $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k$ are **linearly independent**
  - $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k = \mathbf{0} \Rightarrow c_1 = \cdots = c_k = 0$
- Let S_1, S_2 be finite subsets of $\mathbb{R}^n$ such that $S_1 \subseteq S_2$
  - S_1 linearly dependent $\Rightarrow S_2$ linearly dependent
  - S_2 linearly independent $\Rightarrow S_1$ linearly independent
- If $\mathbf{0} \in S$, then S is linearly dependent
- **Theorem** - Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\} \subseteq \mathbb{R}^n, k \geq 2$. Then
  - S is linearly dependent $\iff \exists\ v_i$ st it is a linear combination of the other vectors
  - S is linearly independent $\iff$ no vector in S is the linear combination of other vectors
- **Theorem** - Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\} \subseteq \mathbb{R}^n$. If $k > n$, then S is linearly dependent
- **Theorem** - Suppose $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\} \subseteq \mathbb{R}^n$ is linearly independent. If $\boldsymbol{v}_{k+1}$ is not in $\text{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$, then $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k, \boldsymbol{v}_{k+1}\}$ is linearly independent

3.5 Bases

- A set V is called a **vector space** if V is a subspace of $\mathbb{R}^n$ for some n
- If W and V are vector spaces such that $W \subseteq V$, then W is a **subspace** of V
- Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$ be a subset if a vector space V. Then S is called a **basis** for V if S is linearly independent and $\text{span}(S) = V$
- A basis for a vector space V contains the smallest possible number of vectors that spans V and the largest possible number of vectors that are linearly independent
- For convenience, $\emptyset$ is said to be the basis for $\{\mathbf{0}\}$
- **Theorem** - Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$ be a subset of vector space V. Then S is a basis for V iff every vector $\boldsymbol{v} \in V$ can be uniquely expressed as

$$\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k,\ c_i \in \mathbb{R}$$

- $c_1, c_2, \dots, c_k$ are the **coordinates** of $\boldsymbol{v}$ **relative** to S
- $(c_1, c_2, \dots, c_k)$ is the **coordinate vector** of $\boldsymbol{v}$ **relative** to the basis S, denoted by $(\boldsymbol{v})_S$

* The order of $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k$ above is fixed

- Let $E = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots \boldsymbol{e}_n\}$ be a subset of $\mathbb{R}^n$ with $\boldsymbol{e}_1 = (1, 0, \dots, 0), \boldsymbol{e}_2 = (0, 1, \dots, 0), \boldsymbol{e}_n = (0, 0, \dots, 1)$. Then E is a basis and is called the **standard basis** for $\mathbb{R}^n$
- **Theorem** - Let S be a basis for a vector space V
  - $(\boldsymbol{v})_S = \mathbf{0} \iff \boldsymbol{v} = \mathbf{0}$
  - For any $c \in \mathbb{R}$ and $\boldsymbol{v} \in V$, $(c\boldsymbol{v})_S = c(\boldsymbol{v})_S$
  - For any $\boldsymbol{u}, \boldsymbol{v} \in V$, $(\boldsymbol{u} + \boldsymbol{v})_S = (\boldsymbol{u})_S + (\boldsymbol{v})_S$
  - For any $\boldsymbol{u}, \boldsymbol{v} \in V$, $\boldsymbol{u} = \boldsymbol{v} \iff (\boldsymbol{u})_S = (\boldsymbol{v})_S$
  - For any $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_r \in V$, $c_1, c_2, \dots, c_r \in \mathbb{R}$, $(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_r\boldsymbol{v}_r)_S = c_1(\boldsymbol{v}_1)_S + c_2(\boldsymbol{v}_2)_S + \cdots + c_r(\boldsymbol{v}_r)_S$
- **Theorem** - Let S be a basis for a vector space V. Suppose $|S| = k$. Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_r \in V$
  - $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_r$ are linearly independent $\iff (\boldsymbol{v}_1)_S, (\boldsymbol{v}_2)_S, \dots, (\boldsymbol{v}_r)_S$ are linearly independent
  - $\text{span}(S) = V \iff \text{span}\{(\boldsymbol{v}_1)_S, (\boldsymbol{v}_2)_S, \dots, (\boldsymbol{v}_r)_S\} = \mathbb{R}^k$

3.6 Dimensions

- Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$ be a subset of $\mathbb{R}^n$
  - If $k > n$ then S is linearly dependent
  - If $k < n$, then $\text{span}(S) \neq \mathbb{R}^n$
- **Theorem** - Let V be a vector space having a basis with k vectors
  - Any subset of V with $> k$ vectors is linearly dependent
  - Any subset of V with $< k$ vectors cannot span V

  Sketch of proof: Consider a basis S and express all vectors in terms of coordinates wrt S
- **Corollary** - All bases of a vector space have the same size
- Let V be a vector space and S be a basis for V. Then the **dimension** of V is $\dim(V) = |S|$
- Let $\boldsymbol{Ax} = \boldsymbol{0}$ be a homogeneous linear system whose solution set is the vector space V and $\boldsymbol{R}$ be an REF of $\boldsymbol{A}$. Then
$$\text{no. of non-pivot col in } \boldsymbol{R} = \text{no. of arbitrary parameters in sol} = \text{the dimension of } \boldsymbol{V}$$
- **Theorem** - Let S be a subset of a vector space V. The following are equivalent:
  1. S is a basis for V
  2. S is linearly independent and $|S| = \dim(V)$
  3. S spans V and $|S| = \dim(V)$

  Sketch of Proof: $1 \Rightarrow 2$ and $1 \Rightarrow 3$ are trivial. $2 \Rightarrow 1$ and $3 \Rightarrow 1$ by contradiction
- **Theorem** - Let U be a subspace of a vector space V. Then $U = V \iff \dim(U) = \dim(V)$
  Sketch of Proof: $\Rightarrow$ is trivial. For $\Leftarrow$, show that a basis of U is a basis of V
- **Theorem** - Let $\boldsymbol{A}$ be a square matrix of order n. Then the following are equivalent
  1. $\boldsymbol{A}$ is invertible
  2. $\boldsymbol{Ax} = \boldsymbol{b}$ has a unique solution $\forall\ \boldsymbol{b}$
  3. $\boldsymbol{Ax} = \boldsymbol{0}$ has only the trivial solution
  4. The reduced row-echelon form of $\boldsymbol{A}$ is $\boldsymbol{I}_n$
  5. $\boldsymbol{A}$ is a product of elementary matrices
  6. $\det(\boldsymbol{A}) \neq 0$
  7. The rows of $\boldsymbol{A}$ form a basis for $\mathbb{R}^n$
  8. The columns of $\boldsymbol{A}$ form a basis for $\mathbb{R}^n$

  Sketch of Proof: 1 to 6 are already shown. Convert 7 and 8 to $\boldsymbol{A}$ or $\boldsymbol{A}^T$ is invertible

3.7 Transition Matrices

- Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$ be a basis for a vector space V and $(\boldsymbol{v})_S = (c_1, c_2, \dots, c_k)$ be the coordinate vector of $\boldsymbol{v} \in V$ relative to S
- View each $\boldsymbol{v}_i$ as a column vector. Then $\begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_k \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} = \boldsymbol{v}$
- The column vector $[\boldsymbol{v}]_S = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$ is also called the **coordinate vector** of $\boldsymbol{v}$ relative to S

- Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_k \end{pmatrix}$ Then $\boldsymbol{A}[\boldsymbol{v}]_S = \boldsymbol{v} \ \forall \ \boldsymbol{v} \in V$
- Let $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k\}, T = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$ be bases for a vector space V
- Suppose vectors are all viewed as column vectors and $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_k \end{pmatrix}$, $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_k \end{pmatrix}$
- Let $w \in \boldsymbol{V}$. Then
$$\begin{aligned} \boldsymbol{w} = \boldsymbol{A}[\boldsymbol{w}]_S &= \begin{pmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_k \end{pmatrix} [\boldsymbol{w}]_S \\ &= \begin{pmatrix} \boldsymbol{B}[\boldsymbol{u}_1]_T & \dots & \boldsymbol{B}[\boldsymbol{u}_k]_T \end{pmatrix} [\boldsymbol{w}]_S \\ &= \boldsymbol{B} \begin{pmatrix} [\boldsymbol{u}_1]_T & \dots & [\boldsymbol{u}_k]_T \end{pmatrix} [\boldsymbol{w}]_S \\ \implies & \begin{pmatrix} [\boldsymbol{u}_1]_T & \dots & [\boldsymbol{u}_k]_T \end{pmatrix} [\boldsymbol{w}]_S = [\boldsymbol{w}]_T \end{aligned}$$
- Let V be a vector space and $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots \boldsymbol{u}_k\}$, T be bases for V. Then
$$\boldsymbol{P} = \begin{pmatrix} [\boldsymbol{u}_1]_T & \dots & [\boldsymbol{u}_k]_T \end{pmatrix}$$
is the **transition matrix** from S to T and
$$\boldsymbol{P}[\boldsymbol{w}]_S = [\boldsymbol{w}]_T \ \forall \ w \in V$$
- **Theorem** - Let S and T be bases for a vector space $V$and $\boldsymbol{P}$ be the transition matrix from S to T. Then $\boldsymbol{P}$ is invertible and $\boldsymbol{P}^{-1}$ is the transition matrix from T to S
Sketch of Proof: Let $\boldsymbol{Q}$ be the trans mat from T to S, then $\boldsymbol{QPI} = \boldsymbol{QP} \begin{pmatrix} e_1 & \dots & e_k \end{pmatrix} = \begin{pmatrix} e_1 & \dots & e_k \end{pmatrix} = \boldsymbol{I}$

4 Vector Spaces Associated with Matrices

4.1 Row Spaces and Column Spaces

- Let $\mathbf{A} = (a_{ij})_{m \times n}$
- The **row space** of $\mathbf{A}$ is $\text{span}\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_m\}$, where $\mathbf{r}_i = \begin{pmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{pmatrix}$ is the ith row of $\mathbf{A}$
- The **column space** of $\mathbf{A}$ is $\text{span}\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n\}$, where $\mathbf{c}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$ is the jth column of $\mathbf{A}$
- Row space of $\mathbf{A}$ = Column space of $\mathbf{A}^T$
Column space of $\mathbf{A}$ = Row space of $\mathbf{A}^T$
- **Theorem** - Let $\mathbf{A}$ and $\mathbf{B}$ be matrices of the same size. If $\mathbf{A}$ and $\mathbf{B}$ are row equivalent, then they have the same row space. (In particular, a REF of $\mathbf{A}$ gives a basis for its row space)
Sketch of Proof: Show that all 3 possible EROs don't change the row space
- **Theorem** - Let $\mathbf{A}$ and $\mathbf{B}$ be row equivalent matrices. Then
  - If $\exists$ a linear relation among some columns of $\mathbf{A}$, then it also holds for corresponding columns of $\mathbf{B}$
  - If a set of columns of $\mathbf{A}$ is linearly ind., then the corresponding set of columns of $\mathbf{B}$ is also linearly ind.
  - If a set of columns of $\mathbf{A}$ is a basis for $\mathbf{A}$'s column space, then the corresponding set of columns of $\mathbf{B}$ is also a basis for $\mathbf{B}$'s column space
- There are two ways to **find a basis of a vector space** $V = \text{span}(S)$
  1. View each $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m \in S$ as a row vector, find an REF $\mathbf{R}$ of $\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_m \end{pmatrix}$, and take the nonzero rows of $\mathbf{R}$
  2. View each $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m \in S$ as a column vector, find an REF $\mathbf{R}'$ of $\begin{pmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_m \end{pmatrix}$, find the pivot columns of $\mathbf{R}$, and take the corresponding columns from $\mathbf{V}$
- **Theorem** - Let $\mathbf{A}$ be an $m \times n$ matrix. Then
  - The column space of $\mathbf{A}$ is $\{\mathbf{A}v \mid v \in \mathbb{R}^n\}$
  - The linear system $\mathbf{Ax} = \mathbf{b}$ is consistent $\iff$ $\mathbf{b}$ lies in the column space of $\mathbf{A}$

4.2 Ranks

- **Theorem** - Let $\mathbf{A}$ be a matrix. Then dim(row space of $\mathbf{A}$) = dim(column space of $\mathbf{A}$)
- Let $\mathbf{A}$ be a matrix. The dimension of the row/column space of $\mathbf{A}$ is the **rank** of $\mathbf{A}$ and is denoted by $\text{rank}(\mathbf{A})$
- **Properties of Rank** Let $\mathbf{A}$ be an $m \times n$ matrix
  - $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T)$
  - $\text{rank}(\mathbf{A}) = 0 \iff \mathbf{A} = \mathbf{0}$
  - $\text{rank}(\mathbf{A}) \leq \min\{m, n\}$
  - * $\mathbf{A}$ is called **full rank** if $\text{rank}(\mathbf{A}) = \min\{m, n\}$
  - A square matrix $\mathbf{A}$ is of full rank $\iff$ $\mathbf{A}$ is invertible
- Let $\mathbf{Ax} = \mathbf{b}$ be a linear system and $\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ be the columns of $\mathbf{A}$. Then

$$\mathbf{Ax} = \mathbf{b} \text{ is consistent}$$

$$\iff \mathbf{b} \in \text{span}\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$$

$$\iff \text{span}\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\} = \text{span}\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n, \mathbf{b}\}$$

$$\iff \dim \text{span}\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\} = \dim \text{span}\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n, \mathbf{b}\}$$

$$\iff \text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A} \mid \mathbf{b})$$

  Remark - $\text{rank}(\mathbf{A}) \leq \text{rank}(\mathbf{A} \mid \mathbf{b}) \leq \text{rank}(\mathbf{A}) + 1$

- **Theorem** - Let $\mathbf{A}$ and $\mathbf{B}$ be $m \times n$ and $n \times p$ matrices. Then
  - Column space of $\mathbf{AB}$ $\subseteq$ Column space of $\mathbf{A}$
  - Row space of $\mathbf{AB}$ $\subseteq$ Row space of $\mathbf{B}$

  In particular, $\text{rank}(\mathbf{AB}) \leq \min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\}$
  Sketch of Proof: 1st statement by definition, 2nd statement follows from considering their transpose

4.3 Nullspaces and Nullities

- Let $\mathbf{A}$ be an $m \times n$ matrix. The **nullspace** of $\mathbf{A}$ is the solution space of $\mathbf{Ax} = \mathbf{0}$, i.e. $\{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{0}\}$
- The dimension of the nullspace is called the **nullity** of $\mathbf{A}$, denoted by $\text{nullity}(\mathbf{A})$

* From now on, unless otherwise stated, vectors in the nullspace are viewed as column vectors

- **Theorem** - Let $\mathbf{A}$ be an $m \times n$ matrix. Then

$$\text{rank}(\mathbf{A}) + \text{nullity}(\mathbf{A}) = n$$

  Sketch of Proof: Consider a REF of $\mathbf{A}$ and look at the number of pivot/non-pivot columns

- **Dimension Theorem** - Suppose $\mathbf{Ax} = \mathbf{b}$ has a solution $\mathbf{v}$. Then the solution set of $\mathbf{Ax} = \mathbf{b}$ is

$$\{\mathbf{v} + \mathbf{w} \mid \mathbf{w} \in \text{nullspace of } \mathbf{A}\}$$

5 Orthogonality

5.1 The Dot Product

- Let $\mathbf{u} = (u_1, \dots, u_n), \mathbf{v} = (v_1, \dots, v_n) \in \mathbb{R}$
- The dot product (inner product) of $\mathbf{u}$ and $\mathbf{v}$ is

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \dots u_n v_n$$

- The norm (length) of $\mathbf{v}$ is

$$\|\mathbf{v}\| = \sqrt{v_1^2 + \dots + v_n^2}$$

- $\mathbf{v}$ is called a unit vector if $||\mathbf{v}|| = 1$
- The distance between $\mathbf{u}$ and $\mathbf{v}$ is

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2}$$

- The angle between $\mathbf{u}$ and $\mathbf{v}$ ($\mathbf{u}, \mathbf{v} \neq \mathbf{0}$) is

$$\theta = \cos^{-1}\left(\frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}\right),\ 0^\circ \leq \theta \leq 180^\circ$$

- If $\mathbf{u}$ and $\mathbf{v}$ are viewed as row/column vectors, then $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}\mathbf{v}^T / \mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}$
- **Theorem - Properties of the Dot Product**
  Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $c \in \mathbb{R}$
  1. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
  2. $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
     $\mathbf{w} \cdot (\mathbf{u} + \mathbf{v}) = \mathbf{w} \cdot \mathbf{u} + \mathbf{w} \cdot \mathbf{v}$
  3. $(c\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (c\mathbf{v}) = c(\mathbf{u} \cdot \mathbf{v})$
  4. $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$
  5. $\mathbf{v} \cdot \mathbf{v} \geq 0$ and $\mathbf{v} \cdot \mathbf{v} = 0 \iff \mathbf{v} = \mathbf{0}$
- **Theorem (Related Ineqs)** - Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$
  - $|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\|\|\mathbf{v}\|$ (Cauchy-Schwarz Inequality)
  - $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ (Triangle Inequality)
  - $d(\mathbf{u}, \mathbf{w}) \leq d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w})$ (Triangle Inequality)

5.2 Orthogonal and Orthonormal Bases

- Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. They are said to be orthogonal if $\mathbf{u} \cdot \mathbf{v} = 0$, denoted by $\mathbf{u} \perp \mathbf{v}$
- Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\} \subset \mathbf{R}^n$
  - S is called orthogonal if $\mathbf{v}_i \cdot \mathbf{v}_j = 0\ \forall\ i \neq j$
  - S is called orthonormal if S is orthogonal and every vector in S is a unit vector
- The process of converting an orthogonal set of nonzero vectors to an orthonormal one, $\mathbf{u}_i \to \dfrac{\mathbf{u}_i}{\|\mathbf{u}_i\|}$, is called normalizing
- **Theorem** - Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be an orthogonal set of nonzero vectors in $\mathbb{R}^n$. Then S is linearly independent. (Corollary: True for orthonormal sets too)
  Sketch of Proof: Consider a linear combination of $\mathbf{0}$ and dot product it with $\mathbf{v}_i$ over all i

- Let S be a basis for a vector space
  - S is an **orthogonal basis** if it is orthogonal
  - S is an **orthonormal basis** if it is orthonormal
- **Theorem** - Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ be an orthogonal basis for a vector space V. For any $\mathbf{w} \in V$,

$$(\mathbf{w})_S = \left(\frac{\mathbf{w}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}, \dots, \frac{\mathbf{w}\cdot\mathbf{u}_k}{\mathbf{u}_k\cdot\mathbf{u}_k}\right)$$

$$\mathbf{w} = \left(\frac{\mathbf{w}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\right)\mathbf{u}_1 + \cdots + \left(\frac{\mathbf{w}\cdot\mathbf{u}_k}{\mathbf{u}_k\cdot\mathbf{u}_k}\right)\mathbf{u}_k$$

  Sketch of Proof: Take the dot product of the linear combination of $\mathbf{w}$ and $\mathbf{u}_i$ over all i
- Let $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ (viewed as column vectors) be a subset of $\mathbb{R}^n$ and $\mathbf{A} = (\mathbf{v}_1 \dots \mathbf{v}_k)$. Then
  - $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is orthogonal $\iff$ $\mathbf{A}^T\mathbf{A}$ is diagonal
  - $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is orthonormal $\iff$ $\mathbf{A}^T\mathbf{A} = \mathbf{I}_k$
- **Theorem** - Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be an orthonormal basis for a vector space V. For any $\mathbf{w} \in V$,

$$(\mathbf{w})_S = (\mathbf{w}\cdot\mathbf{v}_1, \dots, \mathbf{w}\cdot\mathbf{v}_k)$$

$$\mathbf{w} = (\mathbf{w}\cdot\mathbf{v}_1)\mathbf{v}_1 + \cdots + (\mathbf{w}\cdot\mathbf{v}_k)\mathbf{v}_k$$

- Let V be a subspace of $\mathbf{R}^n$. $\mathbf{u} \in \mathbf{R}^n$ is **orthogonal (perpendicular)** to V and called the **normal vector** of V if $\mathbf{u}\cdot\mathbf{v} = 0\ \forall\ \mathbf{v} \in V$
- **Theorem** - Let $V = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be a vector space. Then

$$\mathbf{w} \text{ is orthogonal to } V \iff \mathbf{w}\cdot\mathbf{v}_i = 0\ \forall\ i = 1, \dots, k$$

  Sketch of Proof: $\Rightarrow$ by definition, $\Leftarrow$ by considering the coordinate vector of any $\mathbf{v} \in V$
- (Exercise) If W is a subspace of $\mathbb{R}^n$, then $W^\perp = \{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{v} \text{ is orthogonal to } W\}$ is also a subspace of $\mathbb{R}^n$
- Let V be a vector subspace of $\mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^n$. The unique vector $\mathbf{p}$ such that $\mathbf{n} = \mathbf{w} - \mathbf{p}$ is orthogonal to V is called the **projection** of $\mathbf{w}$ onto V
- **Theorem** - Let $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be an orthonormal basis for a vector space V. The projection of $\mathbf{w}$ onto V is

$$(\mathbf{w}\cdot\mathbf{v}_1)\mathbf{v}_1 + (\mathbf{w}\cdot\mathbf{v}_2)\mathbf{v}_2 + \cdots + (\mathbf{w}\cdot\mathbf{v}_k)\mathbf{v}_k$$

  Sketch of Proof: Let $\mathbf{p}$ be the projection and note that both are true iff $\mathbf{w}\cdot\mathbf{v}_i = \mathbf{p}\cdot\mathbf{v}_i\ \forall\ i = 1, \dots, k$
- **Theorem** - Let $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ be an orthogonal basis for a vector space V. The projection of $\mathbf{w}$ onto V is

$$\left(\frac{\mathbf{w}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\right)\mathbf{u}_1 + \left(\frac{\mathbf{w}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\right)\mathbf{u}_2 + \cdots + \left(\frac{\mathbf{w}\cdot\mathbf{u}_k}{\mathbf{u}_k\cdot\mathbf{u}_k}\right)\mathbf{u}_k$$

  which is the sum of the projections of $\mathbf{w}$ onto $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$
- **Gram-Schmidt Process** - Let $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ be a basis for a vector space V. Define

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{u}_1 \\
\mathbf{v}_2 &= \mathbf{u}_2 - \frac{\mathbf{u}_2\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{u}_3 - \frac{\mathbf{u}_3\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{u}_3\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 \\
\vdots\ &\quad \vdots \\
\mathbf{v}_k &= \mathbf{u}_k - \frac{\mathbf{u}_k\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{u}_k\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{u}_k\cdot\mathbf{v}_{k-1}}{\mathbf{v}_{k-1}\cdot\mathbf{v}_{k-1}}\mathbf{v}_{k-1}
\end{aligned}$$

  Then $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is an orthogonal basis for V
  Normalizing with $\mathbf{w}_i = \dfrac{\mathbf{v}_i}{\|\mathbf{v}_i\|}\ \forall\ i = 1, \dots, n$, $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k\}$ is an orthonormal basis for V

- **Theorem** - Let $\mathbf{A}$ be an $m \times n$ matrix whose columns are linearly independent. Then $\exists$ an $m \times n$ matrix $\mathbf{Q}$ whose columns form an orthonormal set and an invertible $n \times n$ upper triangular matrix $\mathbf{R}$ st $\mathbf{A} = \mathbf{QR}$
Sketch of Proof: Form $\mathbf{A}$ with the vectors from the Gram-Schmidt Process and you can make an $\mathbf{R}$ that works.

- When solving $\mathbf{Ax} = \mathbf{b}$, you can write $\mathbf{A} = \mathbf{QR}$ and get

$$(\mathbf{QR})\mathbf{x} = \mathbf{b} \implies \mathbf{Rx} = \mathbf{Q}^T\mathbf{QTx} = \mathbf{Q}^T\mathbf{b}$$

5.3 Best Approximations

- **Theorem** - Let V be a subspace of $\mathbb{R}^n$. For $\mathbf{u} \in \mathbb{R}^n$, let $\mathbf{p}$ be the projection of $\mathbf{u}$ onto V. Then $\mathbf{p}$ is the **best approximation** of $\mathbf{u}$ in V, i.e.

$$d(\mathbf{u}, \mathbf{p}) \leq d(\mathbf{u}, \mathbf{v}) \ \forall \ v \in V$$

and $d(\mathbf{u}, \mathbf{p}) = d(\mathbf{u}, \mathbf{v}) \iff \mathbf{v} = \mathbf{p}$

- Let $\mathbf{A}$ be an $m \times n$ matrix, $\mathbf{b} \in \mathbb{R}^m$. Then $\mathbf{u} \in \mathbb{R}^n$ is a **least squares solution** to the linear system $\mathbf{Ax} = \mathbf{b}$ if

$$\|\mathbf{b} - \mathbf{Au}\| \leq \|\mathbf{b} - \mathbf{Av}\| \ \forall \ \mathbf{v} \in \mathbb{R}^n$$

- **Theorem** - Let $\mathbf{A}$ be an $m \times n$ matrix, $\mathbf{b} \in \mathbb{R}^m$. Let $\mathbf{p}$ be the projection of $\mathbf{b}$ onto the column space of $\mathbf{A}$. Then

$$\|\mathbf{b} - \mathbf{p}\| \leq \|\mathbf{b} - \mathbf{Av}\| \ \forall \ \mathbf{v} \in \mathbb{R}^n$$

i.e. $\mathbf{u}$ is a least squares solution to $\mathbf{Ax} = \mathbf{b} \iff \mathbf{u}$ is a solution to $\mathbf{Ax} = \mathbf{p}$
Sketch of Proof: Remember that the projection is the best approximation of a vector onto a vector space

- **Methodology for finding a least squares solution to $\mathbf{Ax} = \mathbf{b}$**
  1. Find an orthogonal/orthonormal basis for V, the column space of $\mathbf{A}$
  2. Find the projection $\mathbf{p}$ of $\mathbf{b}$ onto V
  3. Solve the linear system $\mathbf{Ax} = \mathbf{p}$ (gives the least squares solution to $\mathbf{Ax} = \mathbf{b}$)

- **Theorem** - $\mathbf{u}$ is a least squares solution to $\mathbf{Ax} = \mathbf{b} \iff \mathbf{u}$ is a solution to $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$
Sketch of Proof:
  - $\mathbf{Au} - \mathbf{b}$ is orthogonal the column space V of $\mathbf{A}$, and thus, every column $\mathbf{a}_i$ of $\mathbf{A}$
  - $\therefore \ \mathbf{a}_i^T(\mathbf{Au} - \mathbf{b}) = 0 \ \forall \ i = 1, \dots, n$ (since dot product is 0)
  - Stack them up to get $\mathbf{A}^T(\mathbf{Au} - \mathbf{b}) = \begin{pmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_n^T \end{pmatrix} (\mathbf{Au} - \mathbf{b}) = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0}$ and rearrange

- **Another method for finding the projection of a vector $\mathbf{b}$ onto a vector space V**
  1. Suppose $V = \text{span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$
  2. Write $\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix}$ ($\mathbf{a}_i$'s viewed as column vectors)
  3. Find a least squares solution $\mathbf{u}$ to $\mathbf{Ax} = \mathbf{b}$ (i.e. a sol to $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$)
  4. The projection $\mathbf{p}$ onto V is $\mathbf{p} = \mathbf{Au}$

5.4 Orthogonal Matrices

- Let $\mathbf{A}$ be a square matrix. $\mathbf{A}$ is an orthogonal matrix if $\mathbf{A}^T\mathbf{A} = \mathbf{I}$

- **Theorem** - Let $\mathbf{A}$ be a square matrix of order n. Then

$$\begin{aligned} \mathbf{A} \text{ is an orthogonal matrix} &\iff \text{columns of } \mathbf{A} \text{ form an orthonormal basis for } \mathbb{R}^n \\ &\iff \text{rows of } \mathbf{A} \text{ form an orthonormal basis for } \mathbb{R}^n \end{aligned}$$

- **Properties of Orthogonal Matrices**

- If $\mathbf{A}$ is an orthogonal matrix, then so is $\mathbf{A}^T$
- If $\mathbf{A}, \mathbf{B}$ are orthogonal matrices of the same size, then so is $\mathbf{AB}$
- For any $m \times n$ matrix $\mathbf{A}$,

$$\mathbf{A}^T\mathbf{A} = \mathbf{I}_n \iff \text{the columns of } \mathbf{A} \text{ form an orthonormal set}$$

$$\mathbf{A}\mathbf{A}^T = \mathbf{I}_m \iff \text{the rows of } \mathbf{A} \text{ for an orthonormal set}$$

- Let $S = \{\mathbf{u}_1, \dots, \mathbf{u}_k\} \subset \mathbb{R}^n$ be an orthonormal set and $\mathbf{P}$ be an orthogonal $n \times n$ matrix. Then $\{\mathbf{P}\mathbf{u}_1, \dots, \mathbf{P}\mathbf{u}_k\}$ is also an orthonormal set
- Let $S = \{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be orthonormal bases for a vector space V and $\mathbf{A} = (\mathbf{u}_1 \ \cdots \ \mathbf{u}_k), \mathbf{B} = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_k)$. Then $\mathbf{P} = \mathbf{B}^T\mathbf{A}$ and $\mathbf{Q} = \mathbf{A}^T\mathbf{B}$ are the transition matrices from S to T and from T to S an are orthogonal matrices

- Let $\mathbf{P}_0 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. Then for any $\mathbf{u} \in \mathbb{R}^2$, $\mathbf{P}_0\mathbf{u}$ is a **rotation** of $\mathbf{u}$ about O by θ anticlockwise
  Sketch of Proof: View $\mathbf{u}$ as a coordinate vector wrt rotated axes then change to standard basis

6 Diagonalization

6.1 Eigenvalues and Eigenvectors

- If $\mathbf{A}$ is a square matrix and $\exists$ invertible matrix $\mathbf{P}$ st $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{D}$ is diagonal, then $\mathbf{A}^m = \mathbf{PD}^m\mathbf{P}^{-1}$
- Let $\mathbf{v}_i$ be the ith column of $\mathbf{P}$ and λ_i be the (i,i)-entry of $\mathbf{D}$. Then $\mathbf{AP} = \mathbf{PD} \implies \mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i$
- Let $\mathbf{A}$ be a square matrix of order n. Suppose that for some $\lambda \in \mathbb{R}$ and nonzero $\mathbf{v} \in \mathbb{R}^n$,

$$\mathbf{Av} = \lambda\mathbf{v}$$

  Then λ is called an eigenvalue of $\mathbf{A}$ and $\mathbf{v}$ is called an eigenvector of $\mathbf{A}$ associated with λ

- Let $\mathbf{A}$ be a square matrix. Then
  - $\det(\lambda\mathbf{I} - \mathbf{A})$ is the characteristic polynomial of $\mathbf{A}$
  - $\det(\lambda\mathbf{I} - \mathbf{A}) = 0$ is the characteristic equation of $\mathbf{A}$
- **Theorem** - Let $\mathbf{A}$ be a square matrix. Then the eigenvalues of $\mathbf{A}$ are precisely all the roots to the characteristic equation $\det(\lambda\mathbf{I} - \mathbf{A}) = 0$
- **Main Theorem for Invertible Matrices** - Let $\mathbf{A}$ be a square matrix of order n. Then the following are equivalent
  1. $\mathbf{A}$ is invertible
  2. The RREF of $\mathbf{A}$ is $\mathbf{I}_n$
  3. The homogeneous linear system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution
  4. The linear system $\mathbf{Ax} = \mathbf{b}$ has only the unique solution $\forall\ \mathbf{b} \in \mathbb{R}^n$
  5. $\mathbf{A}$ is the product of elementary matrices
  6. $\det(\mathbf{A}) \neq 0$
  7. The rows of $\mathbf{A}$ form a basis for $\mathbb{R}^n$
  8. The columns of $\mathbf{A}$ form a basis for $\mathbb{R}^n$
  9. $\operatorname{rank}(\mathbf{A}) = n$
  10. 0 is not an eigenvalue for $\mathbf{A}$
- **Theorem** - Let $\mathbf{A}$ be an upper/lower triangular matrix. Then its eigenvalues are all the diagonal entries of $\mathbf{A}$.
- Let $\mathbf{A}$ be a square matrix and λ be an eigenvalue of $\mathbf{A}$. The eigenspace of $\mathbf{A}$ associated to λ is the nullspace of $\lambda\mathbf{I} - \mathbf{A}$, denoted by E_λ (or $E_{A,\lambda}$), consisting of all the eigenvectors of $\mathbf{A}$ associated to λ and $\mathbf{0}$.
- $\dim E_\lambda \geq 1$ since $\lambda\mathbf{I} - \mathbf{A}$ is singular by definition

6.2 Diagonalization

- Let $\mathbf{A}$ be a square matrix. $\mathbf{A}$ is called **diagonalizable** if there exists an invertible matrix $\mathbf{P}$ such that
$$\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$$
is a diagonal matrix.
- The diagonal entries of $\mathbf{D}$ are the eigenvalues of $\mathbf{A}$ and the columns of $\mathbf{P}$ are eigenvectors of $\mathbf{A}$ associated to these eigenvalues.
- **Theorem** - Let $\mathbf{A}$ be a square matrix of order n. Then $\mathbf{A}$ is diagonalizable $\iff$ $\mathbf{A}$ has n linearly independent eigenvectors.
- Suppose $\det(\lambda\mathbf{I} - \mathbf{A}) = \prod_{i=1}^{k}(\lambda - \lambda_i)^{r_i}$ for distinct λ_i and E_i is the eigenspace of $\mathbf{A}$ associated to λ_i. Then
  - r_i is the **algebraic multiplicity** $a(\lambda_i)$ of λ_i
  - $\dim E_i$ is the **geometric multiplicity** $g(\lambda_i)$ of λ_i
  - $g(\lambda_i) \le a(\lambda_i)$ (proved in MA2101)
  - $\dim E_i < a(\lambda_i)$ for some $i \implies \dim E_1 + \cdots + \dim E_k < n \implies \mathbf{A}$ is not diagonalizable
- **Algorithm of Diagonalization** - Let $\mathbf{A}$ be a square matrix of order n
  1. Solve $\det(\lambda\mathbf{I} - \mathbf{A}) = 0$ to find eigenvalues of $\mathbf{A}$ ($\exists$ complex roots $\implies$ not diagonalizable)
  2. For each eigenvalue λ_i of $\mathbf{A}$, find a basis S_i for the eigenspace E_{λ_i}.
  ($\mathbf{A}$ is diagonalizable $\iff |S_1| + \cdots + |S_k| = n \iff |S_i| = a(\lambda_i)\ \forall\ i$)
  3. If $\mathbf{A}$ is diagonalizable, $S_1 \cup \cdots \cup S_k = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for $\mathbb{R}^n$ and $\mathbf{P} = (\mathbf{v}_1 \cdots \mathbf{v}_n)$ diagonalizes $\mathbf{A}$
- **Theorem** - Let $\mathbf{A}$ be a square matrix of order n. If $\mathbf{A}$ has n distinct eigenvalues, then it is diagonalizable.

6.3 Orthogonal Diagonalization

- A square matrix $\mathbf{A}$ is called **orthogonally diagonalizable** if $\exists$ an orthogonal matrix $\mathbf{P}$ st $\mathbf{P}^T\mathbf{A}\mathbf{P}$ is diagonal. $\mathbf{P}$ is said to orthogonally diagonalize $\mathbf{A}$
- **Theorem** - A square matrix is orthogonally diagonalizable $\iff$ it is a symmetric matrix
- **Algorithm** - Same as above but pick an orthogonal basis for each eigenspace using Gram-Schmidt

6.4 Quadratic Forms and Conic Sections

- A **quadratic form** in n variables $x_1, \ldots, x_n$ is
$$Q(x_1, \ldots, x_n) = \sum_{i=1}^{n} q_{ii}x_i^2 + \sum_{i<j} q_{ij}x_ix_j$$
- Let $\mathbf{x} = (x_1, \ldots, x_n)^T$ and $\mathbf{A} = (a_{ij})_{n\times n}$ be defined by $a_{ii} = q_{ii}$ and $a_{ij} = a_{ji} = \frac{1}{2}q_ij$ for $i < j$. Then
$$Q(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x},\ \mathbf{x} \in \mathbb{R}^n$$
- Since $\mathbf{A}$ is symmetric, it is orthogonally diagonalizable by a matrix $\mathbf{P}$ st
$$\mathbf{P}^T\mathbf{A}\mathbf{P} = \mathbf{D} = \begin{pmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_n \end{pmatrix}$$
Let $\mathbf{y} = \mathbf{P}^T\mathbf{x} = (y_1, \ldots, y_n)^T \in \mathbb{R}^n$. Then $\mathbf{x} = \mathbf{P}\mathbf{y}$ and
$$Q(\mathbf{x}) = (\mathbf{P}\mathbf{y})^T\mathbf{A}(\mathbf{P}\mathbf{y}) = \mathbf{y}^T(\mathbf{P}^T\mathbf{A}\mathbf{P})\mathbf{y} = \mathbf{y}^T\mathbf{D}\mathbf{y} = \lambda_1y_1^2 + \cdots + \lambda_ny_n^2$$

- A **quadratic equation** in variables x and y is $ax^2 + bxy + cy^2 + dx + ey = f$
  The graph of a quadratic equation is a **conic section**
- Degenerate conic sections - whole plane, empty set, a point, a line, a pair of lines
- Non-degenerate conic section - circle, ellipse, hyperbola, parabola
- Standard form of a circle/ellipse
$$\frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1, \quad \alpha, \beta > 0$$
$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \frac{1}{\alpha^2} & 0 \\ 0 & \frac{1}{\beta^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1$$
  Ellipse with major radius $\max\{\alpha, \beta\}$ and minor radius $\min\{\alpha, \beta\}$
- Standard form of a hyperbola
$$\frac{x^2}{\alpha^2} - \frac{y^2}{\beta^2} = 1 \quad \text{or} \quad -\frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1, \quad \alpha, \beta > 0$$
$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \frac{1}{\alpha^2} & 0 \\ 0 & -\frac{1}{\beta^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1 \quad \text{or} \quad \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} -\frac{1}{\alpha^2} & 0 \\ 0 & \frac{1}{\beta^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1$$
  Semi-major axis α or β and semi-minor axis β or α respectively
- Standard form of a parabola
$$x^2 = \alpha y \quad \text{or} \quad y^2 = \alpha x, \quad \alpha \neq 0$$
$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 & -\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0 \quad \text{or} \quad \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -\alpha & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$
  Ellipse with focal length $\frac{|\alpha|}{4} \neq 0$ (distance from vertex to focus)
- To classify $\mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} = f, \ x \in \mathbb{R}^2$
  1. Orthogonally diagonalize $\mathbf{A}$ by $\mathbf{P}^T \mathbf{A} \mathbf{P} = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$
  2. Let $\mathbf{y} = \mathbf{P}^T \mathbf{x}$. Then $\mathbf{y}^T \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} \mathbf{y} + \mathbf{b}^T \mathbf{P} \mathbf{y} = f$
  3. Complete the squares
- Suppose the conic is non-degenerate. Then
  - $\det(\mathbf{A}) > 0 \iff$ ellipse (or circle)
  - $\det(\mathbf{A}) = 0 \iff$ parabola
  - $\det(\mathbf{A}) < 0 \iff$ hyperbola

7 Linear Transformation

7.1 Linear Transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$

- The mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by
$$T\left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}\right) = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \quad \vdots \qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}$$
  is a **linear transformation** from $\mathbb{R}^n$ to $\mathbb{R}^m$
- T is called a **linear operator** on $\mathbb{R}^n$ if $m = n$

- $\mathbf{A} = (a_{ij})_{m\times n}$ satisfies $T(\mathbf{x}) = \mathbf{Ax}\ \forall\ \mathbf{x} \in \mathbb{R}^n$ and is the **standard matrix** for T
- The **identity transformation/identity operator** on $\mathbb{R}^n$ is $I : \mathbb{R}^n \to \mathbb{R}^n$ st $I(\mathbf{x}) = \mathbf{x}\ \forall\ \mathbf{x} \in \mathbb{R}^n$
- The **zero transformation** is $O : \mathbb{R}^n \to \mathbb{R}^n$ st $O(\mathbf{x}) = \mathbf{0}\ \forall\ \mathbf{x} \in \mathbb{R}^n$
- The standard matrix of a linear transformation is unique
- **Theorem** - If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then
  1. $T(c\mathbf{v}) = cT(\mathbf{v})\ \forall\ \mathbf{v} \in \mathbb{R}^n, c \in \mathbb{R}$
  2. $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})\ \forall\ \mathbf{u}, \mathbf{v} \in \mathbb{R}^n$
- Let $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be the standard basis for $\mathbb{R}^n$. Then the standard matrix $\mathbf{A}$ for T is $\begin{pmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n)\end{pmatrix}$
- (General Definition) Let V and W be vector spaces. A mapping $T : V \to W$ is a linear transformation if

$$T(c\mathbf{u} + d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v})\ \forall\ c, d \in \mathbb{R}, \mathbf{u}, \mathbf{v} \in V$$

- Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis for $\mathbb{R}^n$ and $(\mathbf{v})_S = (c_1, \dots, c_n)$ (coordinate vector wrt S). Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. Then

$$T(\mathbf{v}) = \begin{pmatrix} T(\mathbf{v}_1) & \cdots & T(\mathbf{v}_n)\end{pmatrix}\begin{pmatrix} c_1 \\ \vdots \\ c_n\end{pmatrix} = \begin{pmatrix} T(\mathbf{v}_1) & \cdots & T(\mathbf{v}_n)\end{pmatrix}[\mathbf{v}]_S$$

- Let $\mathbf{A}$ be the standard matrix for T, $\mathbf{P} = \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n\end{pmatrix}$, and $\mathbf{B} = \begin{pmatrix} T(\mathbf{v}_1) & \cdots & T(\mathbf{v}_n)\end{pmatrix}$. Then

$$T(\mathbf{v}) = \mathbf{Av} = \mathbf{A}\begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n\end{pmatrix}[\mathbf{v}]_S \implies \mathbf{AP} = \mathbf{B}$$

- Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, $S = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$, $R = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be bases for $\mathbb{R}^n$, $\mathbf{B} = \begin{pmatrix} T(\mathbf{u}_1) & \cdots & T(\mathbf{u})_n\end{pmatrix}$, $\mathbf{C} = \begin{pmatrix} T(\mathbf{v}_1) & \cdots & T(\mathbf{v}_n)\end{pmatrix}$, and $\mathbf{P}$ = transition matrix from S to R. Then $\mathbf{B} = \mathbf{CP}$
- $T : \mathbb{R}^n \to \mathbb{R}^n$ linear operation on $\mathbb{R}^n$ with standard matrix $\mathbf{A}$, $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis for $\mathbb{R}^n$ and $\mathbf{P} = \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n\end{pmatrix}$. Then

$$[T(\mathbf{v})]_S = \mathbf{P}^{-1}\mathbf{AP}[\mathbf{v}]_S$$

  and T can be represented by $[\mathbf{v}]_S \to \mathbf{B}[\mathbf{v}]_S$ where $\mathbf{B} = \mathbf{P}^{-1}\mathbf{AP}$. We say that $\mathbf{A}$ and $\mathbf{B}$ are **similar**
- Let $S : \mathbb{R}^n \to \mathbb{R}^m$, $T : \mathbb{R}^m \to \mathbb{R}^k$ be linear transformations. Let $T \circ S : \mathbb{R}^n \to \mathbb{R}^k$ be the mapping st

$$(T \circ S)(\mathbf{u}) = T(S(\mathbf{u}))\ \forall\ \mathbf{u} \in \mathbb{R}^n$$

  This is called the **composition** of T with S and is also a linear transformation
- If $\mathbf{A}, \mathbf{B}$ are the standard matrices for S, T, then $\mathbf{BA}$ is the standard matrix for $T \circ S$

7.2 Ranges and Kernels

- Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The **range** of T is the set of all images of T

$$\mathrm{R}(T) = \{T(\mathbf{v}) | \mathbf{v} \in \mathbb{R}^n\}$$

- **Theorem** - The range of $T : \mathbb{R}^n \to \mathbb{R}^m$ is given by

$$\mathrm{R}(T) = \mathrm{span}\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)\}$$

  where $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is any basis for $\mathbb{R}^n$
- **Theorem** - The range of linear transformation T with standard matrix $\mathbf{A}$ is the column space of $\mathbf{A}$

- Let T be a linear transformation with standard matrix $\mathbf{A}$. The **rank** of T is the dimension of $\mathrm{R}(T)$

$$\text{rank}(T) = \dim \mathrm{R}(T) = \text{rank}(\mathbf{A})$$

- Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The **kernel** of T is the set of all vectors in $\mathbb{R}^n$ whose image is the zero vector in $\mathbb{R}^m$

$$\text{Ker}(T) = \{\mathbf{v} \in \mathbb{R}^n | T(\mathbf{v}) = \mathbf{0}\}$$

- **Theorem** - The kernel of linear transformation T with standard matrix $\mathbf{A}$ is the nullspace of $\mathbf{A}$
- Let T be a linear transformation with standard matrix $\mathbf{A}$. The **nullity** of T is the dimension of $\text{Ker}(T)$

$$\text{nullity}(T) = \dim \text{Ker}(T) = \text{nullity}(\mathbf{A})$$

- **Dimension Theorem for Linear Transformations** - Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then

$$\text{rank}(T) + \text{nullity}(T) = n$$

7.3 Geometric Linear Transformations

- **Scaling (in $\mathbb{R}^2$)** (can be generalized to higher dimensions cause yes)
  - The standard matrix is $\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$
  - Scaled along x-axis and y-axis by a factor of λ_1 and λ_2 respectively
  - If $\lambda_1 = \lambda_2 = \lambda$, then it's a dilation/contraction if $\lambda > 1/0 < \lambda < 1$
  - Suppose T has standard matrix $\mathbf{A}$ that is diagonalizable with eigenvalues λ_1, λ_2
    - $\exists$ invertible $\mathbf{P} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{pmatrix}$ st $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$
    - T can be viewed as a scaling along the direction of $\mathbf{v}_1$ and $\mathbf{v}_2$ by a factor of λ_1 and λ_2 respectively
- **Reflections (in $\mathbb{R}^2$)**
  - Standard matrix is $\begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix}$
  - Reflect across the line ℓ formed by rotating the x-axis counterclockwise θ degrees (like in trig)
  - Let $\mathbf{n} = \begin{pmatrix} \cos\theta & \sin\theta \end{pmatrix}^T$ be a unit vector on ℓ and $\mathbf{p}$ be the projection of $\mathbf{v}$ on ℓ. Then

$$T(\mathbf{v}) = 2\mathbf{p} - \mathbf{v} = 2(\mathbf{v} \cdot \mathbf{n})\mathbf{n} - \mathbf{v}$$

  - Let $\mathbf{n} = \begin{pmatrix} \sin\theta & -\cos\theta \end{pmatrix}^T$ be a unit vector orthogonal to ℓ and $\mathbf{p}$ be the projection of $\mathbf{v}$ onto $\text{span}\{\mathbf{n}\}$. Then

$$T(\mathbf{v}) = \mathbf{v} - 2\mathbf{p} = \mathbf{v} - 2(\mathbf{v} \cdot \mathbf{n})\mathbf{n}$$

  - (in $\mathbb{R}^3$) Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the reflection wrt the plane $ax + by + cz = 0$ with a, b, c not all 0.. Then $\mathbf{n} = (a, b, c)^T$ is orthogonal to the plane and

$$T(\mathbf{v}) = \mathbf{v} - \left(2\frac{\mathbf{v} \cdot \mathbf{n}}{\|\mathbf{n}\|^2}\right)\mathbf{n}$$

- **Rotations** (slide 84)
  - Standard matrix is $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$
  - Rotation about the origin by θ
  - Every orthogonal matrix with $\det = 1$ is of the above form; if $\det = -1$ it's a reflection + rotation
- **Shears** (slide 90)

- $T\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x + ky \\ y \end{pmatrix}$ is a shear in the x-direction by a factor of k
- $T\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x \\ y + kx \end{pmatrix}$ is a shear in the y-direction by a factor of k
- $T\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x + k_1 z \\ y + k_2 z \\ z \end{pmatrix}$ is a shear in the x-direction by a factor of k_1 and in the y-direction by a factor of k_2

- **Translations (not a linear transformation)**
  - $T\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x + a \\ y + b \end{pmatrix}$ is a translation by $(a, b)^T$
- **2D Computer Graphic**
  - 2D item drawn by connecting $(a_1, b_1), \ldots, (a_n, b_n)$
  - Let $\mathbf{M} = \begin{pmatrix} a_1 & \ldots & a_n \\ b_1 & \ldots & b_n \end{pmatrix}$
  - If T is a scaling/reflection/rotation/translation of $\mathbb{R}^2$ with standard matrix $\mathbf{A}$, the resulting graphic by applying T is $\mathbf{AM}$
  - To translate by $(a, b)^T$, form a **homogenous coordinate system** by mapping $\begin{pmatrix} a \\ b \end{pmatrix} \leftrightarrow \begin{pmatrix} a \\ b \\ 1 \end{pmatrix}$ and do the shear $T\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x + az \\ y + bz \\ z \end{pmatrix}$